

Robust Estimation of Mean Values ^{*}

Xinjia Chen

November 2008

Abstract

In this paper, we develop a computational approach for estimating the mean value of a quantity in the presence of uncertainty. We demonstrate that, under some mild assumptions, the upper and lower bounds of the mean value are efficiently computable via a sample reuse technique, of which the computational complexity is shown to possess a Poisson distribution.

1 Introduction

In many situations, it is desirable to estimate the mean value of a scalar quantity Q which is a function of independent random vectors V and Δ such that the distribution of V is known and that the distribution of Δ is unknown [4]. Namely, it is interested to estimate the expectation of $Q = q(V, \Delta)$, where $q(.,.)$ is a multivariate function. From modeling considerations, it is reasonable to assume that Δ is bounded in norm $\|\cdot\|$, and radially symmetrical and nondecreasing in its probability density function, $f_{\Delta}(\cdot)$ with the following notions:

- (i) The norm, $\|\Delta\|$, of Δ is no greater than a certain value r , i.e., $\|\Delta\| \leq r$;
- (ii) For any realization Δ of Δ , $f_{\Delta}(\Delta)$ depends only on, $\|\Delta\|$, the norm of Δ ;
- (iii) For any Δ_1 and Δ_2 such that $\|\Delta_1\| < \|\Delta_2\|$, $f_{\Delta}(\Delta_1) \geq f_{\Delta}(\Delta_2)$.

Such assumptions have been proposed by Barmish and Lagoa [1] in the context of robustness analysis of control systems, where Δ is referred to as “uncertainty” because of the lack of knowledge of its distribution.

In this paper, we shall focus on the estimation of the expectation $\mathbb{E}[Q] = \mathbb{E}[q(V, \Delta)]$ based on assumptions (i), (ii) and (iii). Such a problem is referred to as *robust estimation* due to the fact that the exact distribution of Δ is not available. In the special case that the maximum norm r of Δ equals 0, the robust estimation problem reduces to a conventional estimation problem. Instead of seeking the exact value of $\mathbb{E}[Q]$ which is obviously impossible, we aim at obtaining upper and lower bounds for $\mathbb{E}[Q]$. It is intuitive that the gap between the upper and lower bounds should be

^{*}The author had been previously working with Louisiana State University at Baton Rouge, LA 70803, USA, and is now with Department of Electrical Engineering, Southern University and A&M College, Baton Rouge, LA 70813, USA; Email: chenxinjia@gmail.com

increasing with respect to r . Since the relation between Q and \mathbf{V} , Δ can be fairly complicated, the Monte Carlo estimation method is the unique and powerful approach.

The remainder of the paper is organized as follows. In Section 2, we derive upper and lower bounds for $\mathbb{E}[Q]$ based on assumptions (i), (ii) and (iii). In Section 3, we propose a Monte Carlo method for the evaluation of the bounds of $\mathbb{E}[Q]$. In particular, we introduce a sample reuse method to substantially reduce the computational complexity. In Section 4, we investigate the computational complexity of the Monte Carlo method implemented with the principle of sample reuse. Section 5 is the conclusion.

2 Bounds of Expectation

In this section, we shall derive upper and lower bounds of $\mathbb{E}[Q] = \mathbb{E}[q(\mathbf{V}, \Delta)]$ based on the assumptions described in Section 1. For this purpose, we have the following fundamental result, which is a slight generalization of the uniform principle proposed by Barmish and Lagoa [1].

Theorem 1 *Let Δ_ρ^u be a random vector with a uniform distribution over $\{\Delta : \|\Delta\| \leq \rho\}$. Define*

$$\mathbb{M}(\rho) = \mathbb{E}[q(\mathbf{V}, \Delta_\rho^u)], \quad \underline{\mathbb{M}}(r) = \inf_{0 < \rho < r} \mathbb{M}(\rho), \quad \overline{\mathbb{M}}(r) = \sup_{0 < \rho < r} \mathbb{M}(\rho).$$

Then, $\underline{\mathbb{M}}(r) < \mathbb{E}[Q] < \overline{\mathbb{M}}(r)$.

See Appendix A for a proof. Theorem 1 reveals that the computation of the bounds of $\mathbb{E}[Q]$ can be reduced to the evaluation of function $\mathbb{M}(\rho)$, which can be accomplished via Monte Carlo simulation. A conventional method is as follows:

Partition interval $(0, r]$ by grid points $r = \rho_1 > \rho_2 > \dots > \rho_m > 0$. Let $\mathcal{B}_\ell = \{\Delta : \|\Delta\| \leq \rho_\ell\}$. For $\ell = 1, \dots, m$, estimate $\mathbb{M}(\rho_\ell)$ as the empirical mean

$$\frac{\sum_{i=1}^N q(\mathbf{V}_i, X_{\ell,i})}{N}$$

where \mathbf{V}_i , $X_{\ell,i}$, $i = 1, \dots, N$ are mutually independent random variables such that $\mathbf{V}_1, \dots, \mathbf{V}_N$ are i.i.d. random samples of \mathbf{V} and $X_{\ell,1}, \dots, X_{\ell,N}$ are i.i.d. random samples uniformly distributed over \mathcal{B}_ℓ . Clearly, the total number of simulations is Nm for estimating $\mathbb{M}(\rho_\ell)$, $\ell = 1, \dots, m$. A major problem with this approach is that the computational complexity can be extremely high, since the number of grid points m is typically a very large number. To overcome such a problem, we shall develop a sample reuse technique in the next section.

3 Sample Reuse

In this section, we shall explore the idea of sample reuse to reduce the computational complexity. The sample reuse method has been proposed by Chen et al. [2, 3] for the robustness analysis

of control systems. The idea of sample reuse is to start simulation from the largest set \mathcal{B}_1 and if it also belongs to smaller subsets the experimental result is saved for later use in the smaller sets. As can be seen from last section, a conventional approach would require a total of Nm simulations. However, due to sample reuse, the actual number of experiments for set \mathcal{B}_ℓ is a random number \mathbf{n}_ℓ , which is usually much less than N . Hence, this strategy saves a significant amount of computational effort.

In order to provide a precise description of the principle of sample reuse, we assume that all random variables are defined in the same probability space $(\Omega, \mathcal{F}, \text{Pr})$. We shall introduce a function \mathcal{G} , referred to as *sample reuse function*, as follows.

Let X_1, \dots, X_m be i.i.d. samples uniformly distributed over \mathcal{A} . Let Y_1, \dots, Y_n be i.i.d. samples uniformly distributed over \mathcal{B} . Let $m \leq n$ and $\mathcal{A} \supset \mathcal{B}$. Define reusable sample size \mathbf{k} such that $\mathbf{k}(\omega)$ is the number of elements of $\{X_i(\omega) \in \mathcal{B} : i = 1, \dots, m\}$ for any $\omega \in \Omega$. Define random variables Z_1, \dots, Z_n such that, for any $\omega \in \Omega$,

$$Z_\ell(\omega) = \begin{cases} X_{i_\ell}(\omega) & \text{for } 1 \leq \ell \leq \mathbf{k}(\omega), \\ Y_\ell(\omega) & \text{for } \mathbf{k}(\omega) < \ell \leq n \end{cases}$$

where i_ℓ , $1 \leq \ell \leq \mathbf{k}(\omega)$ are the indexes of the elements of $\{X_i(\omega) \in \mathcal{B} : i = 1, \dots, m\}$ such that i_ℓ is increasing with respect to ℓ . This process of generating Z_1, \dots, Z_n from X_1, \dots, X_m and Y_1, \dots, Y_n is denoted by

$$(Z_1, \dots, Z_n; \mathbf{k}) = \mathcal{G}(X_1, \dots, X_m; Y_1, \dots, Y_n).$$

With regard to the distribution of Z_1, \dots, Z_n , we have

Theorem 2 *Suppose X_1, \dots, X_m are independent with Y_1, \dots, Y_n . Then, Z_1, \dots, Z_n are i.i.d. samples uniformly distributed over \mathcal{B} .*

See Appendix B for a proof. Now we can use \mathcal{G} to precisely describe the sample reuse algorithm for estimating $\mathbb{M}(\rho_\ell)$, $\ell = 1, \dots, m$. Let $X_{\ell,i}$, $i = 1, \dots, N$ be the random samples uniformly distributed over \mathcal{B}_ℓ for $\ell = 1, \dots, m$. Let $Y_{1,i} = X_{1,i}$ for $i = 1, \dots, N$ and $(Y_{\ell,1}, \dots, Y_{\ell,N}; \mathbf{k}_\ell) = \mathcal{G}(Y_{\ell-1,1}, \dots, Y_{\ell-1,N}; X_{\ell,1}, \dots, X_{\ell,N})$ for $\ell = 2, \dots, m$. As a result of Theorem 1, we have that, for any $\ell \in \{1, \dots, m\}$, random variables $Y_{\ell,i}$, $i = 1, \dots, N$ have the same associated cumulative distribution with that of random variables $X_{\ell,i}$, $i = 1, \dots, N$. This implies that $\frac{1}{N} \sum_{i=1}^N q(\mathbf{V}_i, Y_{\ell,i})$ has the same distribution as that of $\frac{1}{N} \sum_{i=1}^N q(\mathbf{V}_i, X_{\ell,i})$ for $\ell = 1, \dots, m$. Therefore, we can use $\frac{1}{N} \sum_{i=1}^N q(\mathbf{V}_i, Y_{\ell,i})$ as an estimator of $\mathbb{M}(\rho_\ell)$ for $\ell = 1, \dots, m$. By virtue of such sample reuse method, the total number of simulations is reduced from Nm to $N + \sum_{\ell=2}^m \mathbf{n}_\ell$, where $\mathbf{n}_\ell = N - \mathbf{k}_\ell$ for $\ell = 2, \dots, m$. As will be demonstrated in the next section, this can be a huge reduction of complexity for a large m .

4 Poisson Complexity

Since the total number of simulations for using the sample reuse method to estimate $\mathbb{M}(\rho_\ell)$, $\ell = 1, \dots, m$ is $N + \sum_{\ell=2}^m \mathbf{n}_\ell$, it is important to investigate the distribution of $\sum_{\ell=2}^m \mathbf{n}_\ell$. In this regard, we have the following general result.

Theorem 3 *For arbitrary sequence of nested sets $\mathcal{B}_1 \supset \mathcal{B}_2 \supset \dots \supset \mathcal{B}_m$ with $\text{vol}(\mathcal{B}_1) = V_{\max}$ and $\text{vol}(\mathcal{B}_m) = V_{\min}$, the cumulative distribution function of $\sum_{\ell=2}^m \mathbf{n}_\ell$ is bounded from below by the cumulative distribution function of a Poisson random variable \mathbf{P} with mean $\lambda = N \ln \left(\frac{V_{\max}}{V_{\min}} \right)$. That is, $\Pr\{\sum_{\ell=2}^m \mathbf{n}_\ell = 0\} = \Pr\{\mathbf{P} = 0\}$ and $\Pr\{\sum_{\ell=2}^m \mathbf{n}_\ell \leq k\} > \Pr\{\mathbf{P} \leq k\}$ for any positive integer k . Moreover, as the maximum difference of volumes of all consecutive sets tends to be zero, $\sum_{\ell=2}^m \mathbf{n}_\ell$ converges to \mathbf{P} in distribution.*

See Appendix C for a proof. It should be noted that the volume of a set \mathcal{B} , denoted by $\text{vol}(\mathcal{B})$, is referred to the Lebesgue measure of \mathcal{B} in this paper.

As an immediate consequence of Theorem 3, we have

$$\Pr\left\{\sum_{\ell=2}^m \mathbf{n}_\ell > 0\right\} = \Pr\{\mathbf{P} > 0\}, \quad \Pr\left\{\sum_{\ell=2}^m \mathbf{n}_\ell > k\right\} < \Pr\{\mathbf{P} > k\}, \quad k = 1, 2, \dots$$

which implies that

$$\mathbb{E}\left[\sum_{\ell=2}^m \mathbf{n}_\ell\right] = \sum_{k=0}^{\infty} \Pr\left\{\sum_{\ell=2}^m \mathbf{n}_\ell > k\right\} < \sum_{k=0}^{\infty} \Pr\{\mathbf{P} > k\} = \lambda = N \ln \left(\frac{V_{\max}}{V_{\min}} \right).$$

By virtue of Theorem 3, we can derive some simple bounds for the distribution of $\sum_{\ell=2}^m \mathbf{n}_\ell$ as follows.

Theorem 4 $\Pr\{\sum_{\ell=2}^m \mathbf{n}_\ell \geq k\} \leq e^{-\lambda} \left(\frac{\lambda e}{k}\right)^k$ for any number $k > \lambda = N \ln \left(\frac{V_{\max}}{V_{\min}} \right)$. In particular, $\Pr\{\sum_{\ell=2}^m \mathbf{n}_\ell \geq e\lambda\} \leq e^{-\lambda}$ and $\Pr\{X \geq (1 + \epsilon)\lambda\} < \exp\left(-\frac{\epsilon^2 \lambda}{4}\right)$ for $0 < \epsilon < 1$.

See Appendix D for a proof.

Now we apply Theorem 3 to investigate the density of original samples of Δ . Suppose that the volume of $\{\Delta : \|\Delta\| \leq \rho\}$ is proportional to ρ^d where d is the dimension of the set. Let \mathbf{N}_ρ denote the number of original samples included in $\{\Delta : \|\Delta\| \leq \rho\}$ when applying the sample reuse method to interval $[\frac{\rho}{\kappa}, \rho]$. Define the density of samples at radius ρ as $\mathcal{D}(\rho) = \lim_{\delta \rightarrow 0} \frac{\mathbb{E}[\mathbf{N}_{\rho+\delta} - \mathbf{N}_\rho]}{\delta}$. Then, we have the following result.

Theorem 5 $\mathcal{D}(\rho)$ is equal to $\frac{Nd}{\rho} \left(\frac{\kappa\rho}{a}\right)^d$ for $\rho \in (0, \frac{a}{\kappa}]$ and is less than $\frac{Nd}{\rho}$ for $\rho \in (\frac{a}{\kappa}, a]$.

See Appendix E for a proof. From this theorem, we can obtain an upper bound for the expected number of original samples with norm bounded in $[0, a]$. As can be seen from Theorem 5, the density function is unimodal and achieves the largest value at $\rho = \frac{a}{\kappa}$. The density function is displayed by Figure 1.

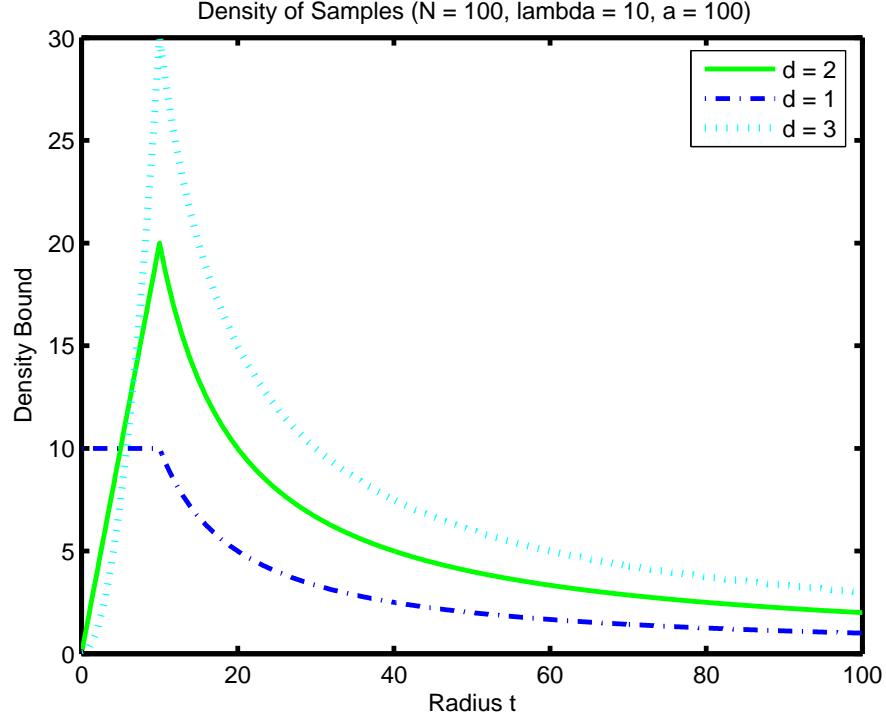


Figure 1: Illustrative Example ($N = 100$, $a = 100$, $\lambda = 10$)

5 Conclusion

We have proposed an efficient computational approach for estimating the mean value of a random function, for which the distribution of relevant random variables are not completely available. A Monte Carlo method with sample reuse as a key mechanism is established. The associated computational complexity is demonstrated to follow a Poisson distribution.

A Proof of Theorem 1

We follow the similar method of Barmish and Lagoa [1]. Let \mathcal{V} denote the volume of $\mathcal{B} = \{\Delta : \|\Delta\| \leq r\}$. We partition the set \mathcal{B} as K layers of equal volume $\frac{\mathcal{V}}{K}$ such that the k -th layer is $\mathcal{L}_k = \{\Delta : r_{k-1} < \|\Delta\| \leq r_k\}$ with $0 = r_0 < r_1 < r_2 < \dots < r_K = r$. Then, the density function can be expressed as

$$f_{\Delta}(\Delta) \approx \sum_{k=1}^K \mathbb{I}_k(\Delta) \lambda_k$$

where λ_k , $k = 1, \dots, K$ satisfying

$$\frac{\mathcal{V}}{K} \sum_{k=1}^K \lambda_k = 1, \quad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_K \geq 0 \quad (1)$$

and $\mathbb{I}_k(\cdot)$ is the indicator function such that $\mathbb{I}_k(\Delta) = 1$ if Δ falls into the k -th layer \mathcal{L}_k and $\mathbb{I}_k(\Delta) = 0$ otherwise. Let $f_{\mathbf{V}}(\cdot)$ denote the density function of \mathbf{V} . Since \mathbf{V} and Δ are independent, we have

$$\begin{aligned}\mathbb{E}[q(\mathbf{V}, \Delta)] &\approx \int_{\{(v, \Delta): \|\Delta\| \leq r\}} q(v, \Delta) f_{\mathbf{V}}(v) dv f_{\Delta}(\Delta) d\Delta \\ &= \int_{\{(v, \Delta): \|\Delta\| \leq r\}} q(v, \Delta) f_{\mathbf{V}}(v) dv \left[\sum_{k=1}^K \mathbb{I}_k(\Delta) \lambda_k \right] d\Delta = \sum_{k=1}^K \alpha_k \lambda_k,\end{aligned}$$

where $\alpha_k = \int_{\{(v, \Delta): \|\Delta\| \leq r\}} q(v, \Delta) \mathbb{I}_k(\Delta) f_{\mathbf{V}}(v) dv d\Delta$. Therefore, the upper and lower bounds of $\mathbb{E}[q(\mathbf{V}, \Delta)]$ correspond to the maximum and minimum of the linear program: $\sum_{k=1}^K \alpha_k \lambda_k$ subject to constraint (1). From convex analysis, the maximum and minimum of this linear program are achieving at extreme points of the form:

$$\lambda_k = \begin{cases} \frac{K}{jV} & \text{for } 1 \leq k \leq j, \\ 0 & \text{for } j < k \leq K. \end{cases}$$

As the number of layers K tends to infinity, the summation $\sum_{k=1}^K \mathbb{I}_k(\Delta) \lambda_k$, which is associated with extreme point $(\lambda_1, \dots, \lambda_K)$, tends to a uniform distribution. This justifies the theorem.

B Proof of Theorem 2

Let $S_\ell \subseteq \mathcal{B}$ for $\ell = 1, \dots, n$. Define $\mathbb{D} = \{1, \dots, n\}$ and $\mathcal{I}_s = \{(i_1, \dots, i_s) : i_1 < \dots < i_s; i_\ell \in \mathbb{D}, \ell = 1, \dots, s\}$. Then,

$$\begin{aligned}\Pr\{Z_\ell \in S_\ell, \ell = 1, \dots, n\} &= \sum_{s=0}^n \sum_{(i_1, \dots, i_s) \in \mathcal{I}_s} \Pr\{X_{i_\ell} \in S_\ell, \ell = 1, \dots, s; X_j \notin \mathcal{B}, j \in \mathbb{D} \setminus \{i_1, \dots, i_s\}\} \\ &\quad \times \Pr\{Y_\ell \in S_\ell, \ell = s+1, \dots, n\}.\end{aligned}$$

For simplicity of notations, we let $V_{S_\ell} = \text{vol}(S_\ell)$, $V_{\mathcal{A}} = \text{vol}(\mathcal{A})$ and $V_{\mathcal{B}} = \text{vol}(\mathcal{B})$. Note that $\Pr\{Y_\ell \in S_\ell, \ell = s+1, \dots, n\} = \prod_{\ell=s+1}^n \left(\frac{V_{S_\ell}}{V_{\mathcal{B}}}\right)$ and

$$\begin{aligned}\Pr\{X_{i_\ell} \in S_\ell, \ell = 1, \dots, s; X_j \notin \mathcal{B}, j \in \mathbb{D} \setminus \{i_1, \dots, i_s\}\} &= \left(\frac{V_{\mathcal{A}} - V_{\mathcal{B}}}{V_{\mathcal{A}}}\right)^{m-s} \prod_{\ell=1}^s \left(\frac{V_{S_\ell}}{V_{\mathcal{A}}}\right) \\ &= \left(\frac{V_{\mathcal{B}}}{V_{\mathcal{A}}}\right)^s \left(1 - \frac{V_{\mathcal{B}}}{V_{\mathcal{A}}}\right)^{m-s} \prod_{\ell=1}^s \left(\frac{V_{S_\ell}}{V_{\mathcal{B}}}\right).\end{aligned}$$

Since there are $\binom{n}{s}$ elements in \mathcal{I}_s , we have

$$\begin{aligned}\Pr\{Z_\ell \in S_\ell, \ell = 1, \dots, n\} &= \sum_{s=0}^n \binom{n}{s} \left(\frac{V_{\mathcal{B}}}{V_{\mathcal{A}}}\right)^s \left(1 - \frac{V_{\mathcal{B}}}{V_{\mathcal{A}}}\right)^{m-s} \prod_{\ell=1}^s \left(\frac{V_{S_\ell}}{V_{\mathcal{B}}}\right) \prod_{\ell=s+1}^n \left(\frac{V_{S_\ell}}{V_{\mathcal{B}}}\right) \\ &= \prod_{\ell=1}^n \left(\frac{V_{S_\ell}}{V_{\mathcal{B}}}\right) \sum_{s=0}^n \binom{n}{s} \left(\frac{V_{\mathcal{B}}}{V_{\mathcal{A}}}\right)^s \left(1 - \frac{V_{\mathcal{B}}}{V_{\mathcal{A}}}\right)^{m-s} \\ &= \prod_{\ell=1}^n \left(\frac{V_{S_\ell}}{V_{\mathcal{B}}}\right).\end{aligned}$$

This concludes the proof of the theorem.

C Proof of Theorem 3

We need some preliminary results.

Lemma 1 *Let $N_1 \leq N_2 \leq \dots \leq N_m$. For $\ell = 1, \dots, m$, let $v_\ell = \text{vol}(\mathcal{B}_\ell)$ and $X_{\ell,i}$, $i = 1, \dots, N_\ell$ be i.i.d. random samples uniformly distributed over \mathcal{B}_ℓ . Let $Y_{1,i} = X_{1,i}$ for $i = 1, \dots, N_1$ and $(Y_{\ell,1}, \dots, Y_{\ell,N_\ell}; \mathbf{k}_\ell) = \mathcal{G}(Y_{\ell-1,1}, \dots, Y_{\ell-1,N_{\ell-1}}; X_{\ell,1}, \dots, X_{\ell,N_\ell})$ for $\ell = 2, \dots, m$. Define $\mathbf{n}_\ell = N_\ell - \mathbf{k}_\ell$ for $\ell = 2, \dots, m$. Then, $\Pr\{\mathbf{n}_\ell = n_\ell, \ell = 2, \dots, m\} = \prod_{\ell=2}^m B\left(N_\ell - n_\ell, N_{\ell-1}, \frac{v_\ell}{v_{\ell-1}}\right)$ for $N_\ell - N_{\ell-1} \leq n_\ell \leq N_\ell$ and $2 \leq \ell \leq m$, where $B(k, n, p) = \binom{n}{k} p^k (1-p)^{n-k}$.*

Proof. We use induction method. First, it is easy to show that the lemma is true for $m = 2$. Next, we assume that the lemma is true for $m - 1$ and show that the lemma is also true for m . Let $\Pr\{(k_1, \dots, k_m), (N_1, \dots, N_m), (v_1, \dots, v_m)\}$ denote the probability that, among the N_1 samples generated from the biggest set \mathcal{B}_1 , there are k_ℓ samples falling into \mathcal{B}_ℓ for $\ell = 1, 2, \dots, m$. Let $\mathbf{P}^m\{(n_2, \dots, n_m), (N_1, \dots, N_m), (v_1, \dots, v_m)\}$ denote the probability of event $\{\mathbf{n}_\ell = n_\ell, \ell = 2, \dots, m\}$ associated with the application of the sample reuse method to sets \mathcal{B}_ℓ , $\ell = 1, \dots, m$ with required sample sizes $N_1 \leq N_2 \leq \dots \leq N_m$. Let $\mathbf{P}^{m-1}\{(n_3, \dots, n_m), (N_2 - k_2, \dots, N_m - k_m), (v_2, \dots, v_m)\}$ denote the probability of event $\{\mathbf{n}_\ell = n_\ell, \ell = 3, \dots, m\}$ associated with the application of the sample reuse method to sets \mathcal{B}_ℓ , $\ell = 2, \dots, m$ with required sample sizes $N_2 - k_2 \leq \dots \leq N_m - k_m$. Note that

$$\begin{aligned} & \mathbf{P}^m\{(n_2, \dots, n_m), (N_1, \dots, N_m), (v_1, \dots, v_m)\} \\ &= \sum_{k_2 \geq k_3 \geq \dots \geq k_m \geq 0} \Pr\{(k_1, \dots, k_m), (N_1, \dots, N_m), (v_1, \dots, v_m)\} \\ & \quad \times \mathbf{P}^{m-1}\{(n_3, \dots, n_m), (N_2 - k_2, N_3 - k_3, \dots, N_m - k_m), (v_2, \dots, v_m)\} \end{aligned}$$

where $n_2 + k_2 = N_2$ and $k_1 = N_1$. By the mechanism of sample reuse,

$$\Pr\{(k_1, \dots, k_m), (N_1, \dots, N_m), (v_1, \dots, v_m)\} = \left[\prod_{\ell=2}^m \binom{k_{\ell-1}}{k_\ell} \left(\frac{v_{\ell-1} - v_\ell}{v_1} \right)^{k_{\ell-1} - k_\ell} \right] \left(\frac{v_m}{v_1} \right)^{k_m}.$$

Since N_ℓ and $-k_\ell$ are non-decreasing with respect to ℓ , we have that $N_\ell - k_\ell$ is non-decreasing with respect to ℓ . Hence, by the assumption of induction,

$$\begin{aligned} & \mathbf{P}^{m-1}\{(n_3, \dots, n_m), (N_2 - k_2, N_3 - k_3, \dots, N_m - k_m), (v_2, \dots, v_m)\} \\ &= \prod_{\ell=3}^m B\left(N_\ell - n_\ell - k_\ell, N_{\ell-1} - k_{\ell-1}, \frac{v_\ell}{v_{\ell-1}}\right) \\ &= \prod_{\ell=3}^m \binom{N_{\ell-1} - k_{\ell-1}}{N_\ell - n_\ell - k_\ell} \left(\frac{v_\ell}{v_{\ell-1}} \right)^{N_\ell - n_\ell - k_\ell} \left(1 - \frac{v_\ell}{v_{\ell-1}} \right)^{N_{\ell-1} - N_\ell + n_\ell - k_{\ell-1} + k_\ell} \end{aligned}$$

and consequently,

$$\begin{aligned}
& \mathbf{P}^m \{(n_2, \dots, n_m), (N_1, \dots, N_m), (v_1, \dots, v_m)\} \\
&= \sum_{k_2 \geq k_3 \geq \dots \geq k_m \geq 0} \left[\prod_{\ell=2}^m \binom{k_{\ell-1}}{k_\ell} \left(\frac{v_{\ell-1} - v_\ell}{v_1} \right)^{k_{\ell-1} - k_\ell} \right] \left(\frac{v_m}{v_1} \right)^{k_m} \\
&\quad \times \prod_{\ell=3}^m \binom{N_{\ell-1} - k_{\ell-1}}{N_\ell - n_\ell - k_\ell} \left(\frac{v_\ell}{v_{\ell-1}} \right)^{N_\ell - n_\ell - k_\ell} \left(1 - \frac{v_\ell}{v_{\ell-1}} \right)^{N_{\ell-1} - N_\ell + n_\ell - k_{\ell-1} + k_\ell} \\
&= \sum_{k_2 \geq k_3 \geq \dots \geq k_m \geq 0} \left[\prod_{\ell=2}^m \binom{k_{\ell-1}}{k_\ell} \binom{N_{\ell-1} - k_{\ell-1}}{N_\ell - n_\ell - k_\ell} \right] \\
&\quad \times \left(\frac{v_m}{v_1} \right)^{k_m} \prod_{\ell=2}^m \left(\frac{v_{\ell-1} - v_\ell}{v_1} \right)^{k_{\ell-1} - k_\ell} \times \prod_{\ell=3}^m \left(\frac{v_\ell}{v_{\ell-1}} \right)^{N_\ell - n_\ell - k_\ell} \left(1 - \frac{v_\ell}{v_{\ell-1}} \right)^{N_{\ell-1} - N_\ell + n_\ell - k_{\ell-1} + k_\ell}.
\end{aligned}$$

Making use of the relationships $k_1 = N_1$ and $k_2 = N_2 - n_2$, we have

$$\begin{aligned}
& \left(\frac{v_m}{v_1} \right)^{k_m} \prod_{\ell=2}^m \left(\frac{v_{\ell-1} - v_\ell}{v_1} \right)^{k_{\ell-1} - k_\ell} \times \prod_{\ell=3}^m \left(\frac{v_\ell}{v_{\ell-1}} \right)^{-k_\ell} \left(1 - \frac{v_\ell}{v_{\ell-1}} \right)^{-k_{\ell-1} + k_\ell} \\
&= (v_1 - v_2)^{k_1 - k_2} \left(\frac{v_2^{k_2}}{v_1^{N_1}} \right) = \left(\frac{v_2}{v_1} \right)^{N_2 - n_2} \left(1 - \frac{v_2}{v_1} \right)^{N_1 - N_2 + n_2}
\end{aligned}$$

and thus

$$\begin{aligned}
& \left(\frac{v_m}{v_1} \right)^{k_m} \prod_{\ell=2}^m \left(\frac{v_{\ell-1} - v_\ell}{v_1} \right)^{k_{\ell-1} - k_\ell} \times \prod_{\ell=3}^m \left(\frac{v_\ell}{v_{\ell-1}} \right)^{N_\ell - n_\ell - k_\ell} \left(1 - \frac{v_\ell}{v_{\ell-1}} \right)^{N_{\ell-1} - N_\ell + n_\ell - k_{\ell-1} + k_\ell} \\
&= \left(\frac{v_m}{v_1} \right)^{k_m} \prod_{\ell=2}^m \left(\frac{v_{\ell-1} - v_\ell}{v_1} \right)^{k_{\ell-1} - k_\ell} \times \prod_{\ell=3}^m \left(\frac{v_\ell}{v_{\ell-1}} \right)^{-k_\ell} \left(1 - \frac{v_\ell}{v_{\ell-1}} \right)^{-k_{\ell-1} + k_\ell} \\
&\quad \times \prod_{\ell=3}^m \left(\frac{v_\ell}{v_{\ell-1}} \right)^{N_\ell - n_\ell} \left(1 - \frac{v_\ell}{v_{\ell-1}} \right)^{N_{\ell-1} - N_\ell + n_\ell} \\
&= \prod_{\ell=2}^m \left(\frac{v_\ell}{v_{\ell-1}} \right)^{N_\ell - n_\ell} \left(1 - \frac{v_\ell}{v_{\ell-1}} \right)^{N_{\ell-1} - N_\ell + n_\ell}.
\end{aligned}$$

On the other hand, if the lemma really holds, we have

$$\begin{aligned}
& \mathbf{P}^m \{(n_2, \dots, n_m), (N_1, \dots, N_m), (v_1, \dots, v_m)\} \\
&= \prod_{\ell=2}^m B \left(N_\ell - n_\ell, N_{\ell-1}, \frac{v_\ell}{v_{\ell-1}} \right) = \left[\prod_{\ell=2}^m \binom{N_{\ell-1}}{N_\ell - n_\ell} \right] \left[\prod_{\ell=2}^m \left(\frac{v_\ell}{v_{\ell-1}} \right)^{N_\ell - n_\ell} \left(1 - \frac{v_\ell}{v_{\ell-1}} \right)^{N_{\ell-1} - N_\ell + n_\ell} \right].
\end{aligned}$$

Therefore, to show the lemma, it remains to show

$$\sum_{k_2 \geq k_3 \geq \dots \geq k_m \geq 0} \left[\prod_{\ell=2}^m \binom{k_{\ell-1}}{k_\ell} \binom{N_{\ell-1} - k_{\ell-1}}{N_\ell - n_\ell - k_\ell} \right] = \prod_{\ell=2}^m \binom{N_{\ell-1}}{N_\ell - n_\ell}.$$

Using the relationships $k_1 = N_1$ and $k_2 = N_2 - n_2$, this identity can be reduced to the following identity

$$\sum_{k_2 \geq k_3 \geq \dots \geq k_m \geq 0} \left[\prod_{\ell=3}^m \binom{k_{\ell-1}}{k_\ell} \binom{N_{\ell-1} - k_{\ell-1}}{N_\ell - n_\ell - k_\ell} \right] = \prod_{\ell=3}^m \binom{N_{\ell-1}}{N_\ell - n_\ell},$$

which can be shown by observing that

$$\begin{aligned}
& \sum_{k_2 \geq k_3 \geq \dots \geq k_{m-i} \geq 0} \left[\prod_{\ell=3}^{m-i} \binom{k_{\ell-1}}{k_{\ell}} \binom{N_{\ell-1} - k_{\ell-1}}{N_{\ell} - n_{\ell} - k_{\ell}} \right] \\
&= \sum_{k_2 \geq k_3 \geq \dots \geq k_{m-i-1} \geq 0} \left[\prod_{\ell=3}^{m-i-1} \binom{k_{\ell-1}}{k_{\ell}} \binom{N_{\ell-1} - k_{\ell-1}}{N_{\ell} - n_{\ell} - k_{\ell}} \right] \sum_{k_{m-i}=0}^{k_{m-i-1}} \binom{k_{m-i-1}}{k_{m-i}} \binom{N_{m-i-1} - k_{m-i-1}}{N_{m-i} - n_{m-i} - k_{m-i}} \\
&= \sum_{k_2 \geq k_3 \geq \dots \geq k_{m-i-1} \geq 0} \left[\prod_{\ell=3}^{m-i-1} \binom{k_{\ell-1}}{k_{\ell}} \binom{N_{\ell-1} - k_{\ell-1}}{N_{\ell} - n_{\ell} - k_{\ell}} \right] \binom{N_{m-i-1}}{N_{m-i} - n_{m-i}}
\end{aligned}$$

for $0 \leq i \leq m-4$ and

$$\sum_{k_2 \geq k_3 \geq 0} \left[\prod_{\ell=3}^3 \binom{k_{\ell-1}}{k_{\ell}} \binom{N_{\ell-1} - k_{\ell-1}}{N_{\ell} - n_{\ell} - k_{\ell}} \right] = \sum_{k_2 \geq k_3 \geq 0} \left[\binom{k_2}{k_3} \binom{N_2 - k_2}{N_3 - n_3 - k_3} \right] = \binom{N_2}{N_3 - n_3}.$$

This completes the proof of the lemma. \square

Lemma 2 Let $\theta > 1$ and $N \geq 1$. Define $L(\theta, k) = \sum_{i=0}^k \binom{N}{i} (1 - \frac{1}{\theta})^i (\frac{1}{\theta})^{N-i}$ and $L_P(\theta, k) = \sum_{i=0}^k \frac{(N \ln \theta)^i}{i!} \exp(-N \ln \theta)$ for $k = 0, 1, \dots, N$. Then, $L(\theta, 0) = L_P(\theta, 0)$ and $L(\theta, k) > L_P(\theta, k)$ for $k = 1, \dots, N$.

Proof. First, it is evident that $L(\theta, 0) = L_P(\theta, 0) = \theta^{-N}$ and $L(\theta, N) = 1 > L_P(\theta, N)$. Hence, it remains to show the lemma for $k = 1, \dots, N-1$. It is easy to show that $\lim_{\theta \rightarrow \infty} L(\theta, k) = \lim_{\theta \rightarrow \infty} L_P(\theta, k) = 0$ and thus $\lim_{\theta \rightarrow \infty} [L(\theta, k) - L_P(\theta, k)] = 0$ for $k = 1, \dots, N-1$. It can also be readily checked that $\lim_{\theta \rightarrow 1} L(\theta, k) = \lim_{\theta \rightarrow 1} L_P(\theta, k) = 1$ and consequently $\lim_{\theta \rightarrow 1} [L(\theta, k) - L_P(\theta, k)] = 0$ for $k = 1, \dots, N-1$. Noting that $\frac{\partial L(\theta, k)}{\partial \theta} = -\frac{N!}{k!(N-k-1)!} (1 - \frac{1}{\theta})^k (\frac{1}{\theta})^{N-k+1}$ and $\frac{\partial L_P(\theta, k)}{\partial \theta} = -\frac{(N \ln \theta)^k}{k!} \frac{N}{\theta^{N+1}}$, we have $\frac{\partial [L(\theta, k) - L_P(\theta, k)]}{\partial \theta} = \frac{N}{k! \theta^{N+1}} \left[(N \ln \theta)^k - \frac{(N-1)! (\theta-1)^k}{(N-k-1)!} \right] > 0$ if and only if $\varphi(\theta) > 0$, where $\varphi(\theta) = \ln \theta - \alpha(\theta-1)$ with $\alpha = \left[\frac{(N-1)!}{(N-k-1)!} \right]^{\frac{1}{k}} \frac{1}{N} < 1$. Since $\varphi(1) = 0$ and $\frac{d\varphi(\theta)}{d\theta} = \frac{1}{\theta} - \alpha$ is positive for $\theta \in (1, \frac{1}{\alpha})$, we have $\varphi(\theta) > 0$ for $\theta \in (1, \frac{1}{\alpha}]$. Since $\varphi(\frac{1}{\alpha}) > 0$, $\lim_{\theta \rightarrow \infty} \varphi(\theta) < 0$ and $\frac{d\varphi(\theta)}{d\theta} < 0$ for $\theta > \frac{1}{\alpha}$, there exists a unique number θ^* greater than $\frac{1}{\alpha}$ such that $\varphi(\theta^*) = 0$. Hence, $\varphi(\theta)$ is positive for $\theta \in (1, \theta^*)$ and negative for $\theta > \theta^*$. This implies that $L(\theta, k) - L_P(\theta, k)$ is monotonically increasing with respect to $\theta \in (1, \theta^*)$ and monotonically decreasing with respect to $\theta \in (\theta^*, \infty)$. Recalling that $\lim_{\theta \rightarrow 1} [L(\theta, k) - L_P(\theta, k)] = \lim_{\theta \rightarrow \infty} [L(\theta, k) - L_P(\theta, k)] = 0$, we have $L(\theta, k) > L_P(\theta, k)$ for any $\theta > 1$. This completes the proof of the lemma. \square

Lemma 3 Let $U_i, V_i, i = 1, \dots, n$ be mutually independent non-negative discrete random variables. Suppose that $\Pr\{U_i = 0\} = \Pr\{V_i = 0\}$ and $\Pr\{U_i \leq k\} > \Pr\{V_i \leq k\}$ for any positive integer k and $i = 1, \dots, n$. Then, $\Pr\{\sum_{i=1}^n U_i = 0\} = \Pr\{\sum_{i=1}^n V_i = 0\}$ and $\Pr\{\sum_{i=1}^n U_i \leq k\} > \Pr\{\sum_{i=1}^n V_i \leq k\}$ for any positive integer k .

Proof. We use induction method. The lemma is obviously true for $n = 1$. Assuming that the lemma is true for $n = m - 1 \geq 1$, we have $\Pr\{\sum_{i=1}^m U_i = 0\} = \Pr\{\sum_{i=1}^{m-1} U_i = 0, U_m = 0\} = \Pr\{\sum_{i=1}^{m-1} V_i = 0\} \Pr\{V_m = 0\} = \Pr\{\sum_{i=1}^m V_i = 0\}$ and $\Pr\{\sum_{i=1}^m U_i \leq k\} = \sum_{l=0}^k \Pr\{\sum_{i=1}^{m-1} U_i = l, U_m \leq k - l\} > \sum_{l=0}^k \Pr\{\sum_{i=1}^{m-1} V_i = l\} \Pr\{V_m \leq k - l\} = \Pr\{\sum_{i=1}^m V_i \leq k\}$ for any positive integer k , which implies that the lemma is also true for $n = m$. By the principle of induction, the lemma is established. \square

We are now in a position to prove the theorem. We shall first show that the distribution of $\sum_{\ell=2}^m \mathbf{n}_\ell$ is bounded from below by the distribution of a Poisson variable with mean $N \ln \frac{V_{\max}}{V_{\min}}$. Define $U_i = \mathbf{n}_{i+1}$ for $i = 1, \dots, m - 1$. Then, by Lemma 1, U_i are independent binomial random variables such that $\Pr\{U_i \leq k\} = L(\theta_i, k)$ for $k = 0, 1, \dots, N$ and $i = 1, \dots, m - 1$, where $\theta_i = \frac{v_i}{v_{i+1}}$. Define Poisson variables V_i , $i = 1, \dots, m - 1$ such that U_i, V_i , $i = 1, \dots, m - 1$ are mutually independent and that $\Pr\{V_i \leq k\} = L_P(\theta_i, k)$ for non-negative integer k and $i = 1, \dots, m - 1$. By Lemmas 2 and 3, we have $\Pr\{\sum_{\ell=2}^m \mathbf{n}_\ell = 0\} = \Pr\{\sum_{i=1}^{m-1} U_i = 0\} = \Pr\{\sum_{i=1}^{m-1} V_i = 0\}$ and $\Pr\{\sum_{\ell=2}^m \mathbf{n}_\ell \leq k\} = \Pr\{\sum_{i=1}^{m-1} U_i \leq k\} > \Pr\{\sum_{i=1}^{m-1} V_i \leq k\}$ for $k = 1, 2, \dots$. Noting that V_1, \dots, V_{m-1} are independent Poisson variables with corresponding means $N \ln \theta_1, \dots, N \ln \theta_{m-1}$, we have that $\sum_{i=1}^{m-1} V_i$ is also a Poisson variable with mean $N \sum_{i=1}^{m-1} \ln \theta_i = N \sum_{i=1}^{m-1} \ln \frac{v_i}{v_{i+1}} = N \ln \frac{v_1}{v_m} = N \ln \frac{V_{\max}}{V_{\min}}$.

Next, we shall show that the distribution of $\sum_{\ell=2}^m \mathbf{n}_\ell$ tends to be the distribution of a Poisson variable with mean $N \ln \frac{V_{\max}}{V_{\min}}$ as $\nu = \max\{v_\ell - v_{\ell+1} : 1 \leq \ell \leq m - 1\}$, the maximum difference between the volumes of two consecutive nested sets, tends to be zero while the volumes of \mathcal{B}_1 and \mathcal{B}_m respectively assume fixed values $v_1 = V_{\max}$ and $v_m = V_{\min}$.

Since all sample sizes are equal to N , by Lemma 1, for $\ell = 2, \dots, m$, the original sample sizes \mathbf{n}_ℓ , $\ell = 2, \dots, m$ are mutually independent binomial random variables such that $\Pr\{\mathbf{n}_\ell = k\} = B(k, N, p_\ell)$ for $0 \leq k \leq N$ and $2 \leq \ell \leq m$, where $p_\ell = 1 - \frac{v_\ell}{v_{\ell-1}}$ with $v_\ell = \text{vol}(\mathcal{B}_\ell)$. Therefore, the moment generating function of $\sum_{\ell=2}^m \mathbf{n}_\ell$ can be expressed as $G(s) = [\prod_{\ell=2}^m (p_\ell s + 1 - p_\ell)]^N$, where $s \in (0, 1]$ is a real number. Since $p_\ell s + 1 - p_\ell$ is positive for any $s \in (0, 1]$ and $\ell = 2, \dots, m$, it is meaningful to define $g(s) = \sum_{\ell=2}^m \ln(p_\ell s + 1 - p_\ell)$ for $s \in (0, 1]$. Hence, $G(s) = \exp(Ng(s))$. For simplicity of notations, define $h(s) = (s - 1) \ln \left(\frac{V_{\max}}{V_{\min}} \right)$, $I_1(s) = \int_0^s \sum_{\ell=2}^m \frac{v_{\ell-1} - v_\ell}{z(v_{\ell-1} - v_\ell) + v_\ell} dz - \int_0^1 \sum_{\ell=2}^m \frac{v_{\ell-1} - v_\ell}{z(v_{\ell-1} - v_\ell) + v_\ell} dz$ and $I_2(s) = \int_0^s \sum_{\ell=2}^m \frac{v_{\ell-1} - v_\ell}{v_\ell} dz - \int_0^1 \sum_{\ell=2}^m \frac{v_{\ell-1} - v_\ell}{v_\ell} dz$. The lemma can be established by the following three steps.

First, it can be seen that $g(s) = I_1(s)$ for any $s \in (0, 1]$, since $I_1(1) = g(1) = 0$ and

$$\frac{dI_1(s)}{ds} = \sum_{\ell=2}^m \frac{v_{\ell-1} - v_\ell}{s(v_{\ell-1} - v_\ell) + v_\ell} = \sum_{\ell=2}^m \frac{p_\ell}{p_\ell s + 1 - p_\ell} = \frac{dg(s)}{ds}$$

for any $s \in (0, 1]$.

Second, we need to show that $|I_1(s) - I_2(s)| \rightarrow 0$ for any $s \in (0, 1]$ as $\nu \rightarrow 0$. Noting that

$$\begin{aligned} \left| \int_0^s \frac{z(v_{\ell-1} - v_\ell)^2}{v_\ell^2 + z(v_{\ell-1} - v_\ell)v_\ell} dz \right| &= \frac{(v_{\ell-1} - v_\ell)^2}{v_\ell} \left| \int_0^s \frac{z}{v_\ell + z(v_{\ell-1} - v_\ell)} dz \right| \\ &\leq \frac{(v_{\ell-1} - v_\ell)^2}{v_\ell} \int_0^s \frac{z}{v_\ell} dz = \frac{s^2(v_{\ell-1} - v_\ell)^2}{2v_\ell^2} \leq \frac{s^2\nu(v_{\ell-1} - v_\ell)}{2V_{\min}^2} \end{aligned}$$

for any $s \in (0, 1]$, we have

$$\begin{aligned} |I_1(s) - I_2(s)| &\leq \sum_{\ell=2}^m \left| \int_0^s \frac{z(v_{\ell-1} - v_\ell)^2}{v_\ell^2 + z(v_{\ell-1} - v_\ell)v_\ell} dz \right| + \sum_{\ell=2}^m \left| \int_0^1 \frac{z(v_{\ell-1} - v_\ell)^2}{v_\ell^2 + z(v_{\ell-1} - v_\ell)v_\ell} dz \right| \\ &\leq \sum_{\ell=2}^m \frac{s^2\nu(v_{\ell-1} - v_\ell)}{2V_{\min}^2} + \sum_{\ell=2}^m \frac{\nu(v_{\ell-1} - v_\ell)}{2V_{\min}^2} \\ &= \frac{(s^2 + 1)\nu}{2V_{\min}^2} \sum_{\ell=2}^m (v_{\ell-1} - v_\ell) = \frac{(s^2 + 1)(V_{\max} - V_{\min})\nu}{2V_{\min}^2}. \end{aligned}$$

Therefore, $|I_1(s) - I_2(s)| \rightarrow 0$ for any $s \in (0, 1]$ and arbitrary v_ℓ , $\ell = 1, \dots, m$, as $\nu \rightarrow 0$.

Third, we need to show $g(s) \rightarrow h(s)$ as $\nu \rightarrow 0$. Since

$$h(s) - I_2(s) = \int_0^s \left[\ln \left(\frac{V_{\max}}{V_{\min}} \right) - \sum_{\ell=2}^m \frac{v_{\ell-1} - v_\ell}{v_\ell} \right] dz - \int_0^1 \left[\ln \left(\frac{V_{\max}}{V_{\min}} \right) - \sum_{\ell=2}^m \frac{v_{\ell-1} - v_\ell}{v_\ell} \right] dz,$$

we have $|I_2(s) - h(s)| \leq \int_0^s \left| \ln \left(\frac{V_{\max}}{V_{\min}} \right) - \sum_{\ell=2}^m \frac{v_{\ell-1} - v_\ell}{v_\ell} \right| dz + \int_0^1 \left| \ln \left(\frac{V_{\max}}{V_{\min}} \right) - \sum_{\ell=2}^m \frac{v_{\ell-1} - v_\ell}{v_\ell} \right| dz$. By the definition of Riemann integration, $\sum_{\ell=2}^m \frac{v_{\ell-1} - v_\ell}{v_\ell} \rightarrow \int_{V_{\min}}^{V_{\max}} \frac{dv}{v} = \ln \left(\frac{V_{\max}}{V_{\min}} \right)$ as $\nu \rightarrow 0$ for arbitrary v_ℓ , $\ell = 1, \dots, m$. It follows that, for any $s \in (0, 1]$ and arbitrary v_ℓ , $\ell = 1, \dots, m$, $|I_2(s) - h(s)| \rightarrow 0$ as $\nu \rightarrow 0$. In view of $|g(s) - h(s)| = |I_2(s) - h(s) + I_1(s) - I_2(s)| \leq |I_2(s) - h(s)| + |I_1(s) - I_2(s)|$, we have $g(s) \rightarrow h(s)$ as $\nu \rightarrow 0$ for any $s \in (0, 1]$ and arbitrary v_ℓ , $\ell = 1, \dots, m$. Therefore, we can conclude that $G(s) \rightarrow \exp \left((s-1)N \ln \left(\frac{V_{\max}}{V_{\min}} \right) \right)$ as $\nu \rightarrow 0$ for any $s \in (0, 1]$ and arbitrary v_ℓ , $\ell = 1, \dots, m$. This proves that $\sum_{\ell=2}^m \mathbf{n}_\ell$ converges in distribution to a Poisson variable of mean $N \ln \left(\frac{V_{\max}}{V_{\min}} \right)$. The proof of the theorem is thus completed.

D Proof of Theorem 4

We need some preliminary results.

Lemma 4 *Let X be a Poisson variable of mean $\lambda > 0$. For any number $k > \lambda$, $\Pr\{X \geq k\} \leq e^{-\lambda} \left(\frac{\lambda e}{k} \right)^k$.*

Proof. Since $\Pr\{X \geq k\} = \Pr\{e^{t(X-k)} \geq 1\} \leq \mathbb{E}[e^{t(X-k)}]$ for any $t > 0$, we have $\Pr\{X \geq k\} \leq \inf_{t>0} \mathbb{E}[e^{t(X-k)}]$. Note that

$$\mathbb{E}[e^{t(X-k)}] = \sum_{i=0}^{\infty} e^{t(i-k)} \frac{\lambda^i}{i!} e^{-\lambda} = e^{\lambda e^t} e^{-\lambda} e^{-tk} \sum_{i=0}^{\infty} \frac{(\lambda e^t)^i}{i!} e^{-\lambda e^t} = e^{-\lambda} e^{\lambda e^t - tk},$$

which is minimized if and only if $\lambda e^t = k$. Since $k > \lambda$, we have $t = \ln\left(\frac{k}{\lambda}\right) > 0$ such that $\lambda e^t = k$. For this value of t , we have $e^{-\lambda} e^{\lambda e^t - tk} = e^{-\lambda} \left(\frac{\lambda e}{k}\right)^k$. Hence, we have shown $\Pr\{X \geq k\} \leq e^{-\lambda} \left(\frac{\lambda e}{k}\right)^k$. \square

Now we are in a position to prove the theorem. By Theorem 3, we have $\Pr\{\sum_{\ell=2}^m \mathbf{n}_\ell \geq k\} \leq \Pr\{X \geq k\} \leq e^{-\lambda} \left(\frac{\lambda e}{k}\right)^k$. Setting $k = e\lambda$, we have $\Pr\{\sum_{\ell=2}^m \mathbf{n}_\ell \geq e\lambda\} \leq e^{-\lambda}$. Moreover, using the inequality $(1 + \epsilon) \ln(1 + \epsilon) > \epsilon + \frac{\epsilon^2}{4}$, $\forall \epsilon \in (0, 1]$, we have $\Pr\{\sum_{\ell=2}^m \mathbf{n}_\ell \geq (1 + \epsilon)\lambda\} < \left[\frac{e^\epsilon}{(1 + \epsilon)^{1 + \epsilon}}\right]^\lambda < \exp\left(-\frac{\epsilon^2 \lambda}{4}\right)$ for $0 < \epsilon < 1$. This completes the proof of the theorem.

E Proof of Theorem 5

By Theorem 3, we have $\mathbb{E}[\mathbf{N}_{\rho+\delta}] \leq N \left[1 + d \ln\left(\frac{\kappa(\rho+\delta)}{a}\right)\right]$. Now fix the gridding over $(\rho, \rho + \delta]$. By Theorem 3, as the gridding over $[\frac{a}{\kappa}, \rho]$ becomes increasingly dense, we have $\mathbb{E}[\mathbf{N}_\rho] \rightarrow N \left[1 + d \ln\left(\frac{\kappa \rho}{a}\right)\right]$. This implies that, for any $\epsilon > 0$, we have $\mathbb{E}[\mathbf{N}_\rho] > N \left[1 + \ln\left(\frac{\kappa \rho}{a}\right)\right] - \epsilon$ for a sufficiently dense gridding over $[\frac{a}{\kappa}, \rho]$. Hence,

$$\mathbb{E}[\mathbf{N}_{\rho+\delta} - \mathbf{N}_\rho] = \mathbb{E}[\mathbf{N}_{\rho+\delta}] - \mathbb{E}[\mathbf{N}_\rho] < Nd \ln\left(\frac{\kappa(\rho+\delta)}{a}\right) - Nd \ln\left(\frac{\kappa \rho}{a}\right) + \epsilon = Nd \ln\left(\frac{\rho+\delta}{\rho}\right) + \epsilon.$$

Since the argument holds for any small $\epsilon > 0$, we have $\mathbb{E}[\mathbf{N}_{\rho+\delta} - \mathbf{N}_\rho] \leq Nd \ln\left(\frac{\rho+\delta}{\rho}\right)$. Therefore, the density $\mathcal{D}(\rho) = \lim_{\delta \rightarrow 0} \frac{\mathbb{E}[\mathbf{N}_{\rho+\delta} - \mathbf{N}_\rho]}{\delta} \leq \lim_{\delta \rightarrow 0} \frac{Nd \ln\left(\frac{\rho+\delta}{\rho}\right)}{\delta} = \frac{Nd}{\rho}$. On the other hand, as the gridding gets dense, we have $\mathbb{E}[\mathbf{N}_{\rho+\delta} - \mathbf{N}_\rho] \rightarrow Nd \ln\left(\frac{\rho+\delta}{\rho}\right)$ and thus $\mathcal{D}(\rho) \rightarrow \frac{Nd}{\rho}$. For $\rho \in (0, \frac{a}{\kappa}]$, it follows from Theorem 3 that \mathbf{N}_ρ is a binomial random variable corresponding to N i.i.d. trials with a success probability $\left(\frac{\kappa \rho}{a}\right)^d$. Hence, $\mathbb{E}[\mathbf{N}_\rho] = N \left(\frac{\kappa \rho}{a}\right)^d$ and accordingly $\mathcal{D}(\rho) = \frac{Nd}{\rho} \left(\frac{\kappa \rho}{a}\right)^d$. This completes the proof of the theorem.

References

- [1] B. R. Barmish and C. M. Lagoa, “The uniform distribution: a rigorous justification for its use in robustness analysis,” *Mathematics of Control, Signals and Systems*, vol. 10, pp. 203–222, 1997.
- [2] X. Chen, K. Zhou, and J. Aravena, “Fast construction of robustness degradation function,” *SIAM Journal on Control and Optimization*, vol. 42, pp. 1960–1971, 2004.
- [3] X. Chen, K. Zhou, and J. Aravena, “Probabilistic robustness analysis — risks, complexity and algorithms,” *SIAM Journal on Control and Optimization*, vol. 47, pp. 2693–2723, 2008.
- [4] P. J. Huber, *Robust Estimation*, Wiley, 1981.